

Onboard Image Registration from Invariant Features

Yi Wang^{1,2}, Justin Ng^{1,2}, Michael J. Garay^{2,3}, Michael C. Burl²

¹California Institute of Technology,
Pasadena, CA 91125
{yiw, junstinn}@caltech.edu

²Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109
{Michael.J.Garay, Michael.C.Burl}@jpl.nasa.gov

³Raytheon
Pasadena, CA

Abstract

This paper describes a feature-based image registration technique that is potentially well-suited for onboard deployment. The overall goal is to provide a fast, robust method for dynamically combining observations from multiple platforms into sensor webs that respond quickly to short-lived events and provide rich observations of objects that evolve in space and time. The approach, which has enjoyed considerable success in mainstream computer vision applications, uses invariant SIFT descriptors extracted at image interest points together with the RANSAC algorithm to robustly estimate transformation parameters that relate one image to another. Experimental results for two satellite image registration tasks are presented: (1) automatic registration of images from the MODIS instrument on Terra to the MODIS instrument on Aqua and (2) automatic stabilization of a multi-day sequence of GOES-West images collected during the October 2007 Southern California wildfires.

1 Introduction

The current suite of spaceborne and in-situ assets (e.g., as deployed and operated by NASA, NOAA, and other groups) provides distributed sensing of the Earth's atmosphere, oceans, and land masses. As part of a project supported through NASA's Earth Science Technology Office (ESTO), we are developing techniques to enable these assets to be dynamically combined into *sensor webs*. A key problem, however, is to precisely relate the observations made by one instrument to the observations made by another instrument. Since many of the observations take the form of images, a fast robust method for achieving automatic image registration is crucial.

The problem of image registration has, of course, been extensively studied. See [22] for a survey or [11] for a recent tutorial. Of this vast body of work, we have tried to focus on techniques that are better suited for sensor web applications, where

there are often additional constraints such as:

- limited computational cycles and RAM
- need for low latency
- distributed data locations (the data for two images may reside on different satellites with a high cost to share data between them)
- robustness to partial occlusion and presence of non-rigid motions between some regions of the images due to cloud cover and cloud motion

Given these constraints, we have avoided some techniques that previously have been used successfully with satellite images such as [19] and [4] as these require dense correlation calculations.

Instead, we have focused on sparse feature-based registration, where the key questions are (1) which features to use and (2) whether to represent the features with just their locations or to attach additional attributes that can aid in matching. As shown by [18], good features to track or to use for registration should have strong gradients in (at least) two distinct directions in a neighborhood of the feature point. Both corner-like and blob-like features satisfy this criteria and have been used extensively, e.g., [12, 7]. Recently, the trend has been to use more complex features with additional attributes. Mikolajczyk and Schmid [17] have evaluated a number of these so-called local descriptor approaches and found that the scale invariant feature transform (SIFT) descriptors developed by David Lowe [15] are quite robust.

Thus, the approach we have adopted for image registration consists of the following steps: (1) extract SIFT descriptors at interest points located at scale-space extrema of the difference-of-gaussians (DOG) operator as in [16], (2) match SIFT descriptors between images to establish tentative correspondences, (3) use RANSAC [6] with the tentative correspondences to estimate the transformation from one image to another and to identify inliers (points that obey the estimated transformation), (4) estimate the epipolar geometry [8, 13] to

refine the set of inliers, and (5) [optionally] use Thin Plate Spline (TPS) [5, 2, 20] to estimate a non-rigid transformation between the images. Similar approaches have been used successfully by a number of researchers for a variety of applications including: Brown and Lowe [3] for stitching together photographic panoramas, Fotin [9] for brain image registration, Lin and colleagues [14] for mosaicking UAV image sequences, and Yin [21] for satellite image registration.

The remainder of the paper is organized as follows. Section 2 describes the approach in more detail. Section 3 presents results for registration of images from two different polar-orbiting satellites, as well as registration of a sequence of images taken from a geostationary satellite. Section 4 provides conclusions.

2 Approach

2.1 SIFT Descriptors

The first step in the approach is to extract SIFT descriptors at interest points in the two images. Here, we closely follow the approach of Lowe [16]. Given an image $I(x, y)$, a difference-of-gaussians operator is applied to the image at multiple scales yielding:

$$D_\sigma(x, y) \triangleq (G_{k\sigma}(x, y) - G_\sigma(x, y)) * I(x, y) \quad (1)$$

where G_σ is a circular 2D Gaussian filter with standard deviation σ . Local extrema of $D_\sigma(x, y)$ are identified to subpixel accuracy. Each interest point is assigned a scale based on σ and an orientation θ based on the local gradient information around the point¹.

SIFT descriptors are extracted from a neighborhood around each interest point, where the neighborhood is defined by the scale and orientation of the interest point. The descriptors (128 values for each interest point) provide a coarse characterization of the gradient orientations in the neighborhood.

2.2 Tentative Correspondences

SIFT descriptors from one image are matched to the SIFT descriptors from another image to establish tentative correspondences. For the experiments reported in Section 3, we simply computed Euclidean distance between the 128-dimensional descriptors and reported two descriptors as matching if they were mutual nearest neighbors and the ratio of the distance from the keypoint to the closet

neighbor to the distance from the keypoint to the second-closest neighbor ≤ 0.8 . The computational complexity of this approach is $O(N_1 \cdot N_2)$, where N_i is the number of features in image i . This brute-force approach can be replaced with more efficient indexing schemes such as the approximate nearest neighbor method developed by Beis and Lowe [1]. In addition, if an initial approximate registration function is available (for example, as recovered from ephemeris and pointing information), the spatial locations of the SIFT descriptors can be used to significantly prune the number of potential matches that must be evaluated.

2.3 RANSAC

Given the tentative correspondences, the next step is to estimate geometrical transformation parameters from one image to the other. Although matching SIFT descriptors has significantly reduced the number of false matches, there are still many false correspondences remaining. RANSAC is widely used in such cases to estimate transformation parameters in the presence of outliers. The main idea is to randomly choose minimal sets of correspondences from among the tentative correspondences, use this minimal set to estimate transformation parameters, and then determine how well the estimated transformation works for the entire set of points. Points (or correspondences) for which the estimated transformation works well are labeled as inliers. Once the inliers have been determined the transformation can be re-estimated using all of the (reliable) available data, which usually improves the result over the transformation estimated with only a minimal set. The method can be used for various transformation classes including translation, affine, homography, and fundamental matrices. The minimal sets for these cases consist of one, three, four, and seven points respectively, although for the case of fundamental matrices we use eight-point sets, which provide more stable estimates in the presence of noise.

The random selection of a minimal set must be repeated for a number of trials to insure that at least one of the randomly selected minimal sets will consist only of valid correspondences. The number of trials, n_{trials} , necessary depends on the size of the minimal set m , the fraction of tentative correspondences that are truly valid, α , and the desired probability p (set by the user) that RANSAC will find at least one outlier-free minimal set.

$$n_{\text{trials}} \geq \frac{\log(1 - p)}{\log(1 - \alpha^m)} \quad (2)$$

¹If an interest point suggests several possible orientations, it is duplicated with each copy corresponding to a different choice of orientation.

For estimating a homography matrix, which requires a minimal set consisting of four points, we typically used a few hundred trials in our experiments.

2.4 Estimation of Epipolar Geometry (Fundamental Matrix)

Epipolar geometry relates two projective views of a scene. Suppose \mathbf{X} is an arbitrary point in 3D space that is imaged by two cameras with centers at \mathbf{C} and \mathbf{C}' , respectively. The line joining \mathbf{C} and \mathbf{C}' is called the baseline. Designate the image of \mathbf{X} on the first image plane to be \mathbf{x} and on the second image plane to be \mathbf{x}' . The epipolar geometry tells that the image \mathbf{x}' , corresponding to \mathbf{x} , must lie on the epipolar line \mathbf{l}' , which is the intersection of the epipolar plane formed by \mathbf{C} , \mathbf{C}' , and \mathbf{x} and the second image plane. The epipolar line is related to \mathbf{x} by the, so-called, fundamental matrix \mathbf{F} . The relation is given by:

$$\mathbf{l}' = \mathbf{F}\mathbf{x} \quad (3)$$

where \mathbf{x} is (3×1) in homogeneous coordinates and \mathbf{F} is (3×3) . The epipolar geometry imposes the constraint that:

$$0 = \mathbf{x}'^T \mathbf{F} \mathbf{x} \quad (4)$$

if \mathbf{x}' and \mathbf{x} are corresponding points. In practice, the constraint will not be exactly satisfied due to noise.

The fundamental matrix \mathbf{F} can, in principle, be estimated from a minimum of 7 correspondences. However, the seven-point algorithm is extremely sensitive, so a more robust eight-point algorithm is used [13]. Here again RANSAC is used to estimate the fundamental matrix. Because most outliers from the initial set of tentative correspondences have been eliminated, a high-quality estimate can be established with a modest number of RANSAC trials.

2.5 Thin Plate Spline Refinement

Optionally, the image registration can be further refined using a non-rigid thin plate spline transformation [20, 2]. This technique has been used extensively in medical and biological image registration applications. TPS is smooth and consists of a global affine transformation together with nonlinear terms that can map a set of K control points to desired positions subject to a particular smoothness constraint. The parameters for a TPS function can be efficiently computed using the QR matrix factorization. The result of using TPS for image registration

on the pair of MODIS-Aqua and MODIS-Terra images is shown in Figure 5.

3 Experimental Results

In this section, we present experimental results for two applications:

- automated registration of images taken by the MODIS instrument on Terra and the MODIS instrument on Aqua taken 105 minutes apart
- stabilization of a multi-day sequence of images from GOES-West captured during the October 2007 severe wildfires in Southern California.

3.1 Terra-Aqua Matchups

Figure 1(a) was taken by the MODIS-Terra satellite at 17:50pm and Figure 1(b) was taken by the MODIS-Aqua satellite at 19:35pm on the same day (they are cropped from the original images to focus on the land and make the size reasonable). The two images are subject to change in viewpoint, rotation, distortion, illumination and slight changes in scale. Looking at the two images closely, one will also notice the deformation of clouds, as well as the appearance and disappearance of some clouds. To facilitate quantitative evaluation of the results, 48 easily recognizable points were hand-labeled in each image to serve as ground-truth. These are shown as numbered points in Figure 1.

Figure 2 shows the DOG interest points extracted from the MODIS-Aqua image. The dots show the location and the arrows show the scale and dominant orientation. Approximately 10^4 interest points were extracted. A similar result was obtained for the MODIS-Terra image.

Figure 3 shows the tentative correspondences between the MODIS-Aqua image and the MODIS-Terra image as obtained by keypoint matching. Corresponding keypoints in the two subplots are labeled with the same integer index. There are around 277 tentative matches, so only 2.7% of the keypoints in one image found a match in the other image. Although this may seem to be a bad result (because such a low fraction of points are matched), the good news is that about 80% of the matches that are found are valid. We can see that there are a large number of tentative matches on the land and mountains and many fewer tentative matches on the lakes and clouds since these areas lack significant texture.

Once the features are extracted and tentatively matched based on their SIFT descriptors, RANSAC is used to estimate an initial geometrical transformation between the images. Features that obey the estimated transformation are labeled as inliers. Figure 4 illustrates the inliers obtained using RANSAC

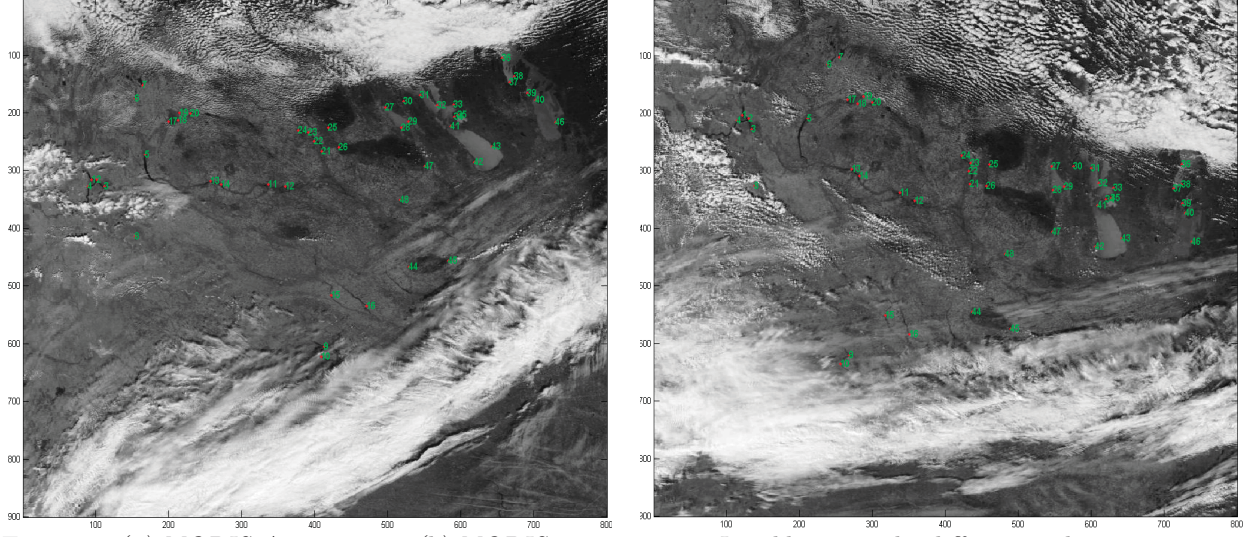


Figure 1: (a) MODIS-Aqua image. (b) MODIS-Terra image. In addition to the differences due to viewpoint, there are also significant differences due to the 105 minute time separation, e.g., due to cloud motion. The numbered points show hand-labeled ground-truth locations of fiducial points used to assess the performance; these are not used in the actual registration process.

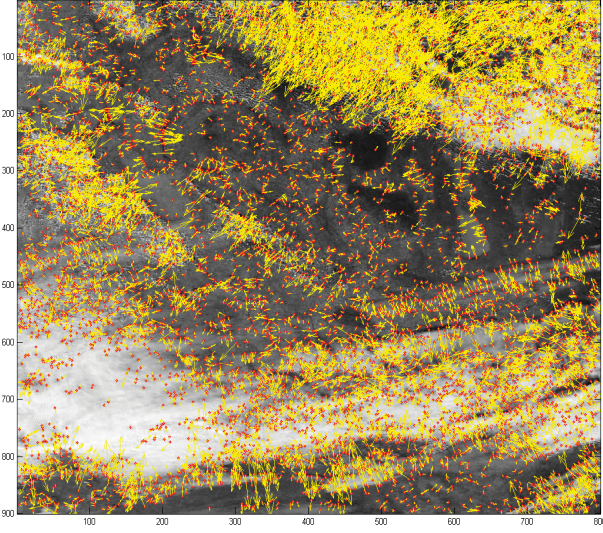


Figure 2: DOG interest points detected on an image from the MODIS-Aqua instrument. The dots show the location and the arrows show the scale and dominant orientation. Note that interest points are generally not found within textureless regions.

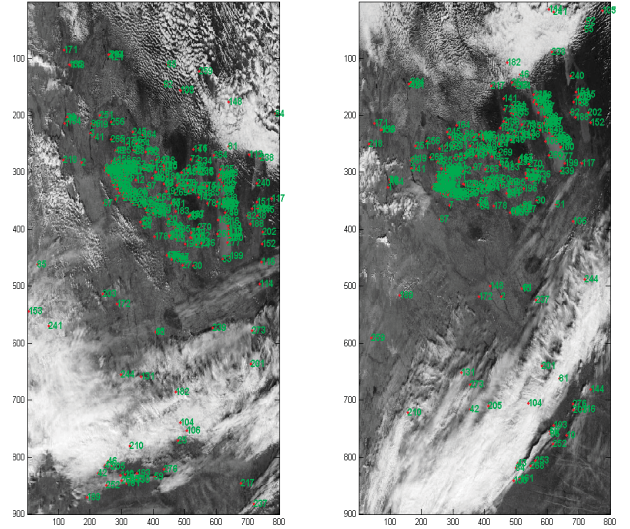


Figure 3: Tentative correspondences based on matching SIFT descriptors. Corresponding points have the same integer-valued label. (a) Aqua. (b) Terra.

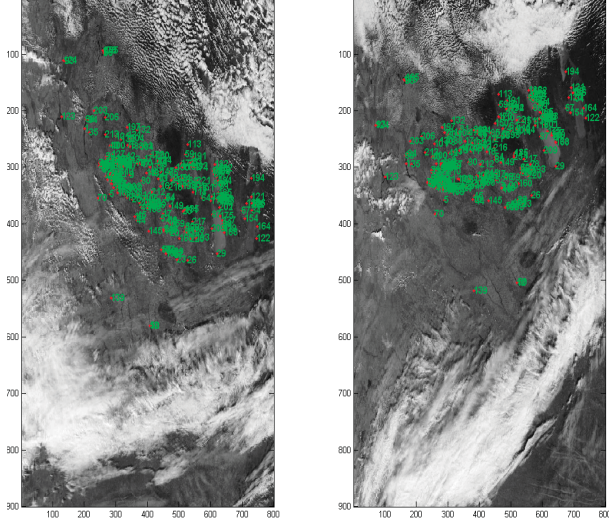


Figure 4: *Inliers after estimation of homography transformation and epipolar geometry using RANSAC. (a) Aqua. (b) Terra. Corresponding points have the same integer-valued label.*

with a homography transformation and refining the inliers based on the estimated epipolar geometry.

Several tests were performed using different classes of geometrical transformation including affine, homography, and TPS. To evaluate the results, we use the ground truth (hand-labeled) fiducial points. The ground-truth points on the MODIS-Terra image are mapped to the MODIS-Aqua image through the estimated transformation. Because RANSAC selects slightly different sets of inliers each time, each experiment was repeated 30 times and the results averaged to measure the performance (we also report the standard deviations to give a feel for the variability). The average distance between the mapped points and the ground truth locations on the MODIS-Aqua image are:

- **affine:** 5.2 pixels with standard deviation 0.41 pixels
- **homography:** 5.4 pixels with a standard deviation of 0.68 pixels
- **affine plus TPS:** 2.6 pixels, with a standard deviation of 0.30 pixels
- **homography plus TPS:** 2.7 pixels, with a standard deviation of 0.38 pixels.

The performance of the affine registration is slightly better than that of homography; however, TPS is the best overall, based on both the quantitative experiments and visual comparisons. The registration result via TPS is shown in Figure 5. One can see that the land is registered well, and the motion of clouds can easily be observed.

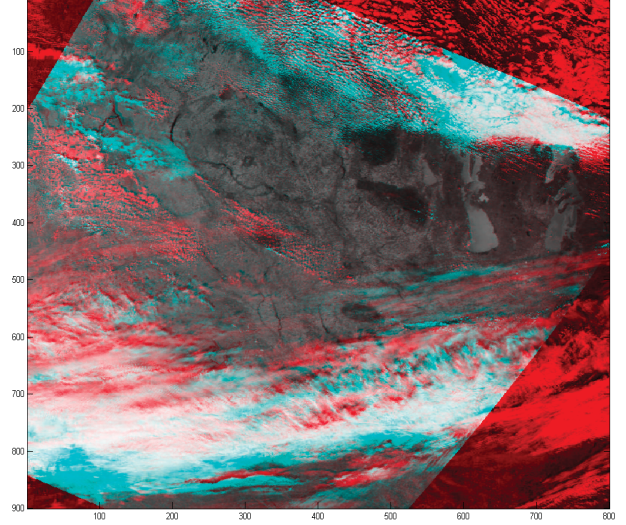


Figure 5: *Registration of the MODIS-Terra image onto the MODIS-Aqua image. The red shows the MODIS-Aqua, the cyan shows the warped MODIS-Terra, and the gray-scale shows where they overlap.*

3.2 GOES Sequence Stabilization

A second application of our technique involves automatically aligning a sequence of GOES West images taken from geostationary orbit during the October 2007 Southern California wildfires. The sequence consists of 124 daytime images covering 4 days. Although GOES is geostationary, there is significant jitter in the raw image sequence due to satellite drift and other errors. To understand the amount of jitter, thirty fiducial points were hand-labeled in each frame². Figure 6 shows the fiducial points on one frame of the sequence. Let the coordinates of the k -th fiducial point in the f -th frame be:

$$\mathbf{p}_k^{(f)} = \begin{pmatrix} x_k^{(f)} \\ y_k^{(f)} \end{pmatrix} \quad (5)$$

Figure 7 illustrates the amount of jitter in the raw sequence. This plot shows how each of the fiducial points would line up if the frames of the raw sequence were simply superimposed on top of each other, i.e., each subplot corresponds to a particular value of k and plots $\mathbf{p}_k^{(f)}$ for each value of f . One can see that all the dots for a given fiducial point fall within a roughly ± 4 pixel window.

To reduce the jitter, we applied the procedure described in Section 2 to automatically register all frames to a designated *base frame* (which happens

²In some frames a few of the fiducial points are not visible due to cloud cover and, in some cases, smoke. These obscured points are omitted from the quantitative analysis.

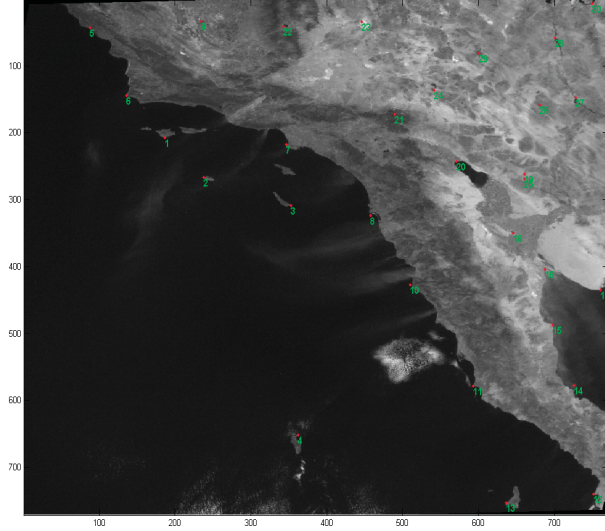


Figure 6: *The designated base frame in the GOES sequence. The hand-labeled fiducial points are also shown.*

to be the frame shown in Figure 6). Note that the hand-labeled fiducial points are not used in the registration; they are only used to quantitatively measure the quality of the final registration. As in the previous experiment, we considered several classes of geometric transformations including translation, affine, homography, and TPS.

Once all instances of fiducial point k are mapped to the base frame via the chosen transformation method, error ellipses based on the covariance matrix (95% confidence region) are used to quantify the performance. Let the coordinates of point k after mapping from frame f to the base frame be designated as:

$$\tilde{\mathbf{p}}_k^{(f)} = (\tilde{x}_k^{(f)}, \tilde{y}_k^{(f)}) \quad (6)$$

Ideally, the mapped versions of point k from all frames would fall precisely on the ground truth location of point k in the base frame. In general, however, this will not happen. The error can be broken up into two components: (1) *bias*: the difference between the mean location of the mapped versions of point k and the ground truth location and (2) *variance*: the spread of the mapped versions of point k relative to the mean location. If desired, the bias (squared) and variance can be combined to measure the mean squared distance of the mapped versions of point k from the ground truth location. In our experiments, we report the area of the 95% confidence ellipse rather than the variance, as the area is somewhat easier to interpret.

Figures 8(a)–(d) show the results based on translation, affine, homography, and TPS, respectively.

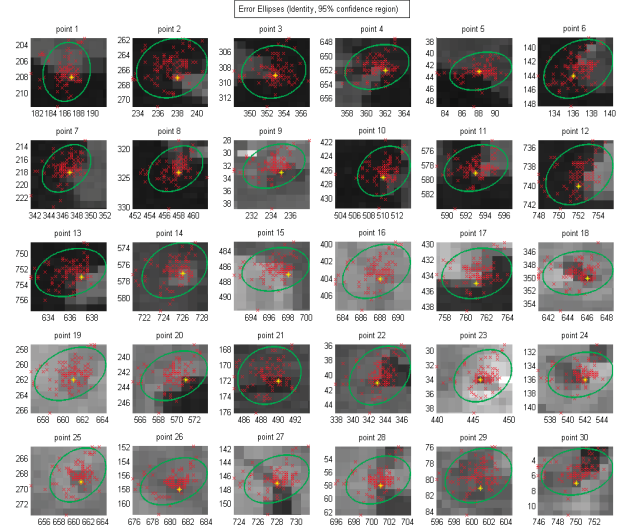


Figure 7: *Error ellipses showing the jitter in the raw GOES sequence. The red crosses are the dots for a given fiducial point mapped to the base frame without any transformation. The yellow asterisk shows the point in the base frame. The green ellipses shows the 95% confidence regions based on the covariance matrix.*

The scales for each subplot are the same so that they can be compared directly. In this case, the simplest transformation, translation, gives the best result. This conclusion is also supported quantitatively by Table 1. Each row in the table corresponds to registration using a different class of transformations. The “identity” transformation on the first row is the same as not doing registration (accepting the raw sequence as registered). The three columns marked *bias* measure the displacement between the mean location of fiducial point k after registration and the location of the same fiducial point in the base frame. The three columns marked *area* measure the area of the 95% confidence ellipse. For both the *bias* and *area* columns, the labels *min*, *med*, and *max* refer to the minimum, median, and maximum values over the 30 points. The translation and affine transformations provide similar results, with translation providing slightly smaller error ellipses. All methods significantly reduce the jitter compared to the raw sequence.

4 Conclusion

The current suite of spaceborne and in-situ sensors provides distributed observations of the Earth’s atmosphere, oceans, and land surface. Onboard image registration techniques will allow these assets to work together as sensor webs enabling ap-

	bias			area		
Transformation	min	med	max	min	med	max
identity	0.2134	1.0374	1.5505	30.9797	38.3572	47.9587
translation	0.0681	0.5409	1.0668	2.6131	7.2531	25.7685
affine	0.0606	0.5315	1.0837	2.6466	7.6042	26.7026
homography	0.1139	0.5144	1.0551	3.1711	9.9174	28.0463
TPS	0.0569	0.5722	0.9766	3.4868	13.7394	38.0297

Table 1: Summary of registration accuracy for 30 fiducial points in the GOES-West image sequence. *Bias* measures the displacement between the mean location of fiducial point k after registration and the location of fiducial point k in the reference frame. *Area* measures the area of the 95% elliptical confidence region.

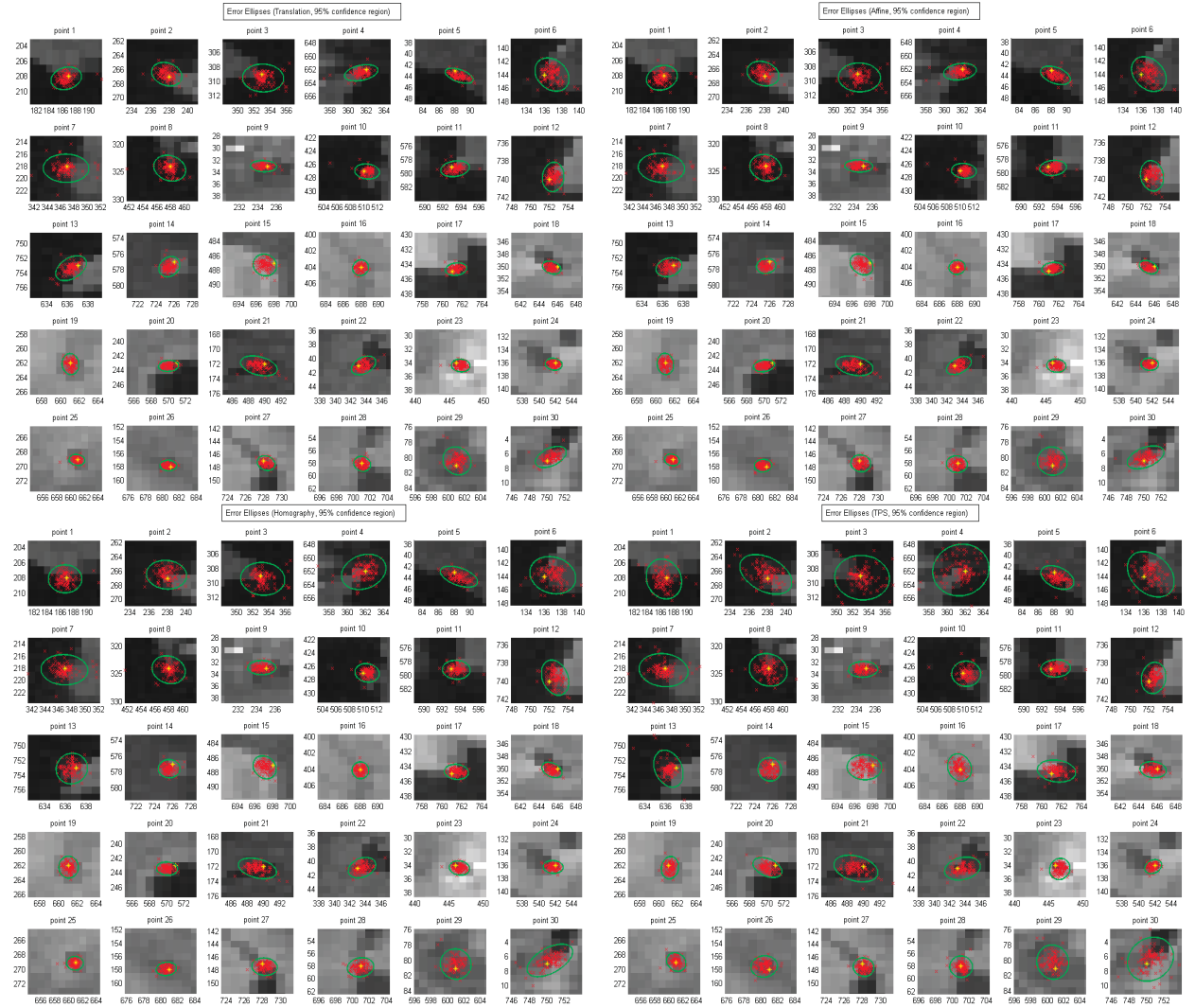


Figure 8: (a) Translation. (b) Affine. (c) Homography, (d) TPS. Each subplot corresponds to a different fiducial point k . The red crosses show $\bar{\mathbf{p}}$, the fiducial points from all frames mapped to the base frame base. The yellow asterisk shows the ground truth position of the fiducial point in the base frame. The green ellipse shows the 95% confidence region on the covariance matrix.

plications such as change detection, feature tracking, and autonomous response to scientifically interesting events. We have presented a potentially viable approach using invariant SIFT descriptors extracted from DOG interest points and RANSAC to perform automatic registration. The method can support various classes of geometric transformations including, translation, affine, homography, epipolar, and TPS. This method does not require geolocations (longitudes and latitudes) of individual pixels or approximate initial registration (although such information can be exploited to speed up the feature matching process).

Acknowledgments

This research has been carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Y.W. and J.N. were supported through Summer Undergraduate Research Fellowships.

References

- [1] J.S. Beis and D.G. Lowe, "Shape indexing using approximate nearest-neighbor search in high-dimensional spaces", In *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 1000–1006, (1997).
- [2] F. L. Bookstein, "Principal Warps: Thin Plate Splines and the Decomposition of Deformations", *IEEE Trans. Pattern Anal. Mach. Intell.*, 11, 567–585, (1989).
- [3] M. Brown and D. G. Lowe, "Recognising Panoramas", In *Proc. 9th Int. Conf. on Computer Vision (ICCV2003)*, pp. 1218–1225, (2003).
- [4] N.A. Bryant, A.L. Zobrist, T.L. Logan, W.L. Bunch, "AFIDS, A Precision Automatic Co-Registration Process for Spacecraft Sensors", *American Geophysical Union, Fall Meeting*, (2004).
- [5] J. Duchon, "Splines minimizing rotation invariant seminorms in sobolev spaces, in *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller (eds), vol. 1, pp. 85–100, Springer-Verlag, (1977).
- [6] M. Fischler, R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Comm. of the ACM*, (1981).
- [7] W. Forstner, E. Gulch, "A Fast Operator for Detection and Precise Location of Distinct Points, Corners, and Centers of Circular Features", *Workshop on Fast Processing of Photogrammetric Data*, pp. 281–305, (1987).
- [8] D. Forsyth, J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, (2003).
- [9] S. Fotin, A. Reeves, "SIFT for Brain Registration", URL: <http://people.cornell.edu/pages/svf3/664project/index.html>
- [10] M.J. Garay, "The Angular Nature of the Shortwave Radiation Reflected from the Earth Observed by GOES-10", M.S. Thesis, University of California, Los Angeles, (2004).
- [11] A. Goshtasby, G. Stockman, K. Rohr, "Workshop on 2-D and 3-D Image Registration", *IEEE Computer Society Conf. of Computer Vision and Pattern Recognition*, Washington D.C., (June 2004).
- [12] C. Harris, M. Stephens, "A combined corner and edge detector", *Proc. of the 4th Alvey Vision Conf.*, pp. 147–151, (1988).
- [13] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, (2004).
- [14] Y. Lin, Q. Yu, G. Medioni, "Map-Enhanced UAV Image Sequence Registration", *Workshop on Application of Comp. Vision (WACV07)*, (2007).
- [15] D. Lowe, "Object Recognition from Local Scale-Invariant Features", *ICCV*, (1999).
- [16] D. Lowe, "Distinctive image features from scale-invariant keypoints", *Int. J. of Computer Vision*, (2004).
- [17] K. Mikolajczyk, C. Schmid, "A Performance Evaluation of Local Descriptors", *Comp. Society Conf. on Computer Vision and Pattern Recognition*, (2003).
- [18] J. Shi, C. Tomasi, "Good Features to Track", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR94)*, (Jun 1994).
- [19] P. Stolorz, R. Blom, R. Crippen, C. Dean, "QUAKEFINDER: Photographing Earthquakes from Space", Technical Report, Machine Learning Systems Group, JPL, (2000). URL: <http://www-aig.jpl.nasa.gov/publications/mls/quakefinder/>
- [20] G. Wahba, "Spline models for observational data", *Society for Industrial and Applied Mathematics (SIAM)*, (1990).
- [21] S. Yin, "Nonrigid Registration of Satellite Images", URL: <http://www.comp.nus.edu.sg/cs6240/past-projects/satellite/satellite.html>
- [22] B. Zitova, J. Flusser, "Image Registration Methods: A Survey", *Image and Vision Computing*, 21(11):977–1000, (2003).